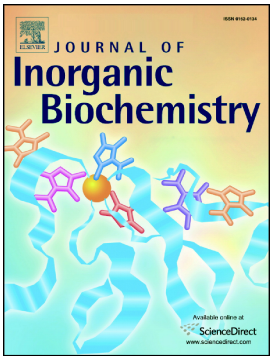


Funneled angle landscapes for helical proteins

John J. Kozak, Harry B. Gray, Roberto A. Garza-López

PII: S0162-0134(20)30119-7

DOI: https://doi.org/10.1016/j.jinorgbio.2020.111091

Reference: JIB 111091

To appear in: *Journal of Inorganic Biochemistry*

Received date: 4 January 2020

Revised date: 14 April 2020

Accepted date: 16 April 2020

Please cite this article as: J.J. Kozak, H.B. Gray and R.A. Garza-López, Funneled angle landscapes for helical proteins, *Journal of Inorganic Biochemistry* (2018), https://doi.org/10.1016/j.jinorgbio.2020.111091

Funneled angle landscapes for helical proteins

John J. Kozak [a], Harry B. Gray [b] and Roberto A. Garza-López [c,*]

a) Department of Chemistry, DePaul University, Chicago IL 60604-6116.
b) Beckman Institute, California Institute of Technology, Pasadena, CA 91125
c) Department of Chemistry and Seaver Chemistry Laboratory, Pomona College, Claremont, CA 91711

Abstract

We use crystallographic data for four helical iron proteins (cytochrome c-b_{562}, cytochrome c', sperm whale myoglobin, human cytoglobin) to calculate radial and angular signatures as each unfolds from the native state stepwise though four unfolded states. From these data we construct an angle phase diagram to display the evolution of each protein from its native state; and, in turn, the phase diagram is used to construct a funneled angle landscape for comparison with the topography of its folding energy landscape. We quantify the departure of individual helical and turning regions from the areal, angular profile of corresponding regions of the native state. This procedure allows us to identify the similarities and differences among individual helical and turning regions in the early stages of unfolding of the four helical heme proteins.

Key words protein folding, cytochromes, myoglobin

Introduction

The groundbreaking theoretical investigations of Wolynes, Saven, and Onuchic [1-7] on funneled energy landscapes together with definitive experimental work by Dobson, Eaton, and others [8-25] have taken our understanding of protein folding to a new level. One aspect that has been of particular interest to us is research by Thirumalai [26-29] using force fields that depend on the radial and angular orientation of sequences of residues to recognize native conformations. Building on this foundational work, we have been developing a geometric model that will allow us to look more deeply into the bottom to top motions in the funneled landscapes of helical proteins. It has long been recognized [30] that when the polypeptide chains and amino acid side chains of proteins are folded into the specific conformations that exist in the native protein, the regions included in the constitutive volumes of different portions of the molecule will fail to pack perfectly with each other. As a result, voids occur in some parts of the folded molecule, and

compressed regions also may appear. The conformation increment is made up of the net contribution of these voids and compressed regions.

Of interest here are the folding energy landscapes of cytochromes b_{562} and c-b_{562}, as they are unlike those of many other cytochromes [11-12]. The unfolded to folded excursions of these four-helix bundle proteins occur with minimal frustration, that is, they avoid misfolded intermediates. Remarkably, the presence of the heme in these proteins does not introduce misligated traps in their energy landscapes [11].

Why are the landscapes of helix bundles so different from those of other cytochromes and globins? In attacking this problem, we built on recent work on the stereochemistry of residues in helical and non-helical regions of proteins [31-32], where two spatial signatures were used to analyze the residue coordinate data; and a third was introduced to analyze amino-acid molecular volume data. The residue-by-residue analysis, taken together with the finding that two major factors stabilize an α-helix (minimization of side-chain steric interference and intrachain H-bonding), led to the conclusion that certain residues are preferentially selected for α-helix formation. In the sequential, de novo synthesis of a turning region, residues are selected such that the overall molecular volume profile (representing purely repulsive, excluded volume effects) spans a small range Δ of values (Δ = 39.1Å) relative to the total range that could be spanned (Δ = 167.7 Å). Thus, excluded volume effects are of enormous importance for residues in helical regions as well as those in adjacent turning regions. Once steric effects are taken into account, down-range attractive interactions between residues come into play in the formation of α-helical regions. The geometry of α-helices can be accommodated by conformational changes in less structured turning regions of a polypeptide, thereby producing a globally optimized (native) protein structure.

In full accord with the findings from work on funneled energy landscapes [1-7], we concluded that the interplay between residues in helical and turning regions drives protein folding. Steric interference between and among side chains is minimized in the first stages of folding, followed by optimization of attractive interactions (formation of H-bonds between specific pairs of amino acids), and finally by torsional adjustment of turning regions to accommodate all α-helical structures.

We have examined the conformational changes accompanying unfolding of helical proteins from the native state to the top of a funneled landscape employing our geometric method of structure analysis [31-32]. We report funneled angle landscapes for cytochrome c-b_{562} (cyt c-b_{562}) [33], cytochrome c', (cyt c') [34], sperm whale myoglobin (sw-Mb) [35] and human cytoglobin (h- Cygb) [36].

Spatial and angular signatures of helical and turning regions

The starting point in our approach is a triplet module of three residues, a center residue (i) flanked by its two first nearest neighbors $(i-1)$ and $(i+1)$. We define a coordinate system in which the iron atom is assigned as the origin. Using crystallographic data for a given protein, we calculate the distance $R(i-1)$ between the iron atom and the α-carbon of the left-most residue, the distance $R(i+1)$ to the right-most residue, and the distance $R(i-1 \text{ to } i+1)$ between the two α-carbons of the terminal residues. Also calculated from crystallographic data are the angles between $R(i-1)$ and $R(i-1 \text{ to } i+1)$, $R(i-1)$ and $R(i+1)$, and $R(i-1 \text{ to } i+1)$ and $R(i+1)$, designated α, β, γ, respectively. These signatures are compiled for each of the n residues of the protein. Analogous calculations have been carried out for sequences of five, seven, eleven and fifteen residues.

Continuing, we next calculate the distance $T(i)$ between the terminal α-carbons $[i-2 \text{ to } i+2]$ for a configuration in which the triplet $[i-2, i-1, i]$ is annexed to the triplet $[i, i+1, i+2]$. This planar configuration may be thought of as an unfolded state, as it is different from the native configuration. By construction, $T(i)$ is greater than or equal to the native state distance, $R(i-2)$ to $R(i+2)$, so that for all residues $i = 2$ to $i = n-1$ we have

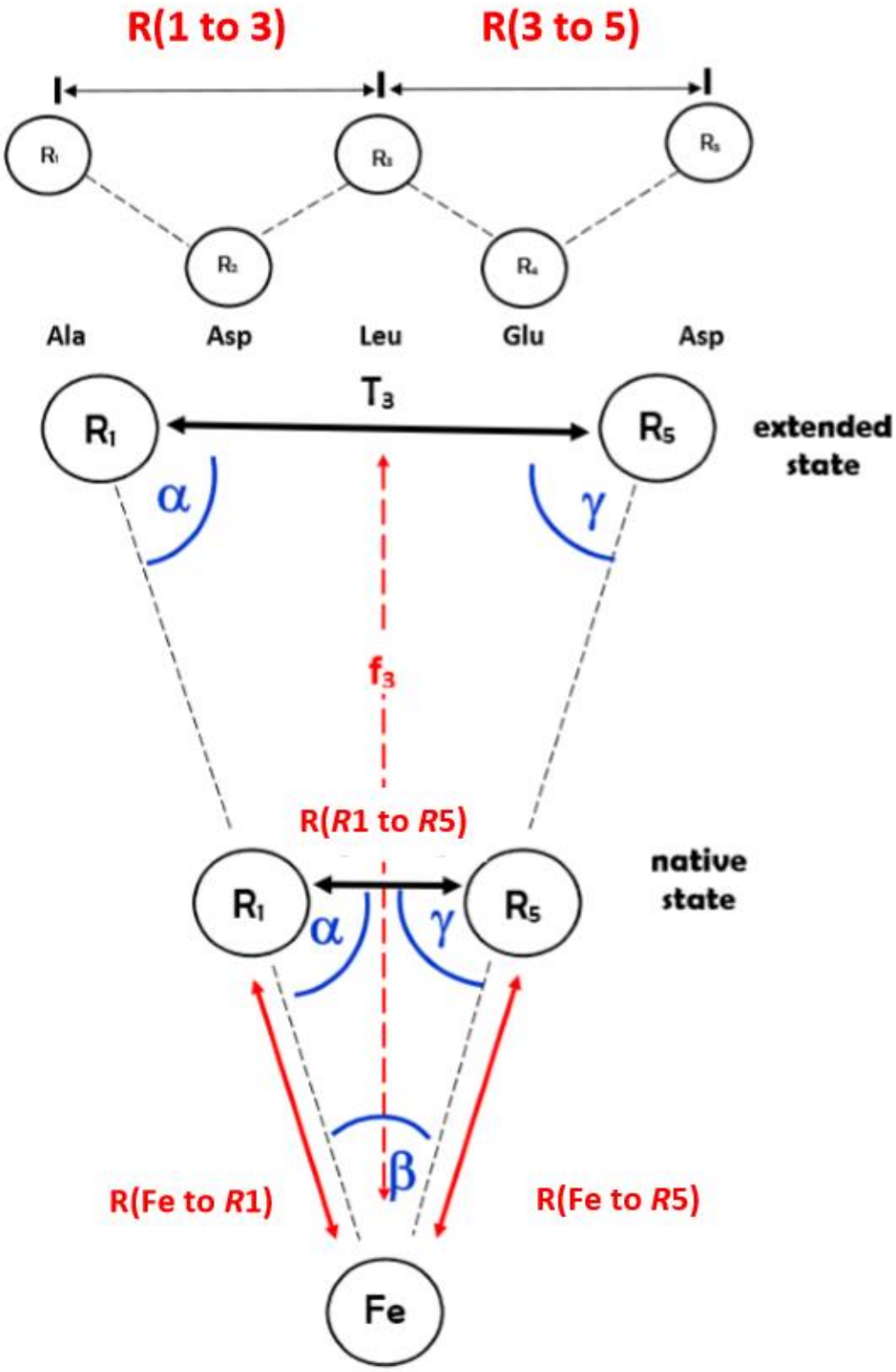
$$\text{Ratio} = \frac{T(i)}{R(i-2) \text{ to } R(i+2)} \geq 1$$

Using the Law of Sines and Cosines, we established in previous work [32] that an exact analytical expression can be derived for the displacement of a central residue in a *n*-residue segment from the iron atom as the protein unfolds. For example, consider the first five residues in cyt c-b_{562} [see Figure 1]. For the five-residue segment centered on residue 3, the displacement f_3 of residue 3 relative to the iron atom in the first stage of unfolding is given by

$$f_3 = \frac{T_3 \sin\big(\alpha(1 \text{ to } 5)\big)}{R1\big[sin\big(\beta(1\ to\ 5)\big)\big]}$$

Angles in calculation of f_3 are in radians.) Expressing $\alpha(1 \text{ to } 5)$ and $\beta(1 \text{ to } 5)$ in terms of the spatial coordinates $R1$, $R5$ and $R(1 \text{ to } 5)$, we get

Figure 1: Specification of the geometrical model for a five-residue segment in cytochrome c-b_{562}.



$$\alpha(1 \text{ to } 5) = \arccos\left[(R5^2 + R(1 \text{ to } 5)^2 - R1^2)\sqrt{2R5\, R(1 \text{ to } 5)}\right]$$

$$= 119.247^0$$

$$\beta(1 \text{ to } 5) = \arccos\left[(R1^2 + R5^2 - R(1 \text{ to } 5)^2)\sqrt{2\, R1\, R5}\ \right]$$

$$= 22.475^0$$

where, T_3 = 12.302, $R1$= 14.175 Å, $R5$ = 10.064 Å and

$R(1 \text{ to } 5)$ = 6.211 Å

Calculation then gives

$$f_3 = \frac{T_3 \sin\left(\arccos((R5^2 + R(1 \text{ to } 5)^2 - R1^2)\sqrt{2R5\, R(1 \text{ to } 5)}\ )\right)}{R1\left[sin\left(\arccos((R1^2 + R5^2 - R(1 \text{ to } 5)^2)\sqrt{2\, R1\, R5}\ )\right)\right]}$$

$= 1.981$ (Angles in calculation of f3 are in radians.)

The expression for f_3 can be re-expressed exactly in an equivalent expression that is useful in interpreting the results obtained from our analysis. The proof $f_3 = \frac{T_3}{R(1 \text{ to } 5)}$ is in Appendix 1.

Note that the analysis is general and equivalence can be established for any of the residues in cyt c-b_{562}, cyt c', sw-Mb and h-Cygb, and for any stage of unfolding. We have confirmed this exact equivalence via direct calculations for all residues and all stages of unfolding for the investigated helical heme proteins.

Spatial signature for unfolding of a polypeptide chain

We established earlier [31] that the elongation of the polypeptide chain as a protein begins to unfold can be quantified. Specifically, the ratio in the first stage of unfolding

$$\text{Ratio}(i) = \frac{T(i)}{R(i-2) \text{ to } R(i+2)}$$

gives the extension of the polypeptide chain when the residues $(i-2)$,$(i-1)$, i,$(i+1)$ and $(i+2)$ are fully extended (two coplanar, adjacent triplets) relative to the native state. Plots of this ratio for cyt c-b_{562} and h-Cygb for the first and sixth stages of unfolding were given previously [31].

The profiles of this ratio versus residue number for the second $T(i)/R(i-3)$ to $R(i+3)$ and fourth, $T(i)/R(i-5)$ to $R(i+5)$, stages of cyt c-b_{562} unfolding are shown in Fig. 2. From Figs. 1 and 5 in [44] and Fig. 2 we can see that the profiles of $T(i)/R(i-3)$ to $R(i+3)$versus residue number and $T(i)/R(i-5)$ to $R(i+5)$ versus residue number are qualitatively and quantitatively quite similar. However, both qualitative and quantitative differences between the first two stages of unfolding and the fourth and sixth stages of unfolding are evident, which may be understood by recalling that there are 3.6 amino acid residues per turn of an α-helix (a five-residue segment would capture one turn). A seven-residue segment would bracket nearly (but not quite) two turns, eleven residues three turns and fifteen residues four turns. Thus, the number of hydrogen bonds broken in the first and second stage of unfolding is the same, but increases stepwise in the fourth and sixth stages of unfolding.

We find this result to be quite general. Displayed in Table 1 are values of the overall average of each ratio calculated for all four heme proteins. The similarity in values at each stage of unfolding for each protein is noteworthy. Again, small quantitative differences characterize the first two stages of unfolding, while more significant differences are apparent in the fourth and sixth stages.

Figure 2. Ratio vs residue number for cyt c-b_{562}: *left*: ratio $T(i)/R(i-3)$ to $R(i+3)$ and *right*: ratio $T(i)/R(i-5)$ to $R(i+5)$

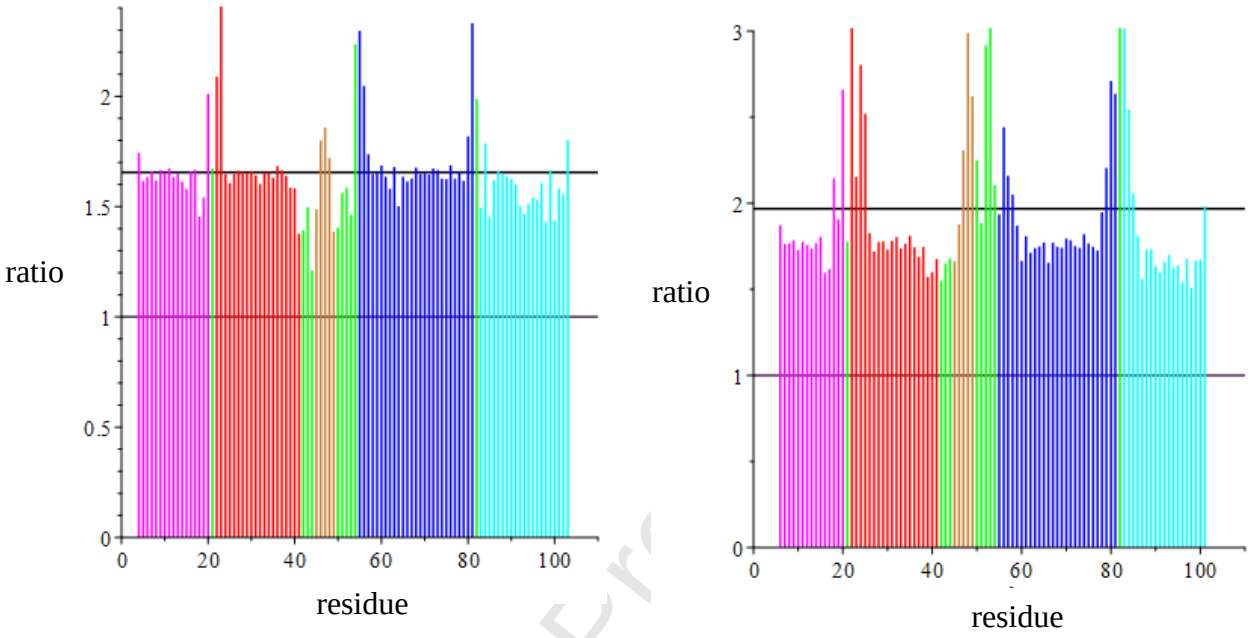


TABLE 1. Average elongation of the protein and individual helical regions. Standard deviation is specified.

Ratio	cyt c-b_{562}	cyt c'	sw-Mb	h-Cygb
$\frac{T(i)}{R(i-1) \text{ to } R(i+1)}$	1.0	1.0	1.0	1.0
$\frac{T(i)}{R(i-2) \text{ to } R(i+2)}$	1.650	1.604	1.674	1.601
H1	1.70 ± 0.21	1.67 ± 0.18	1.67 ± 0.23	1.73 ± 0.23
H2	1.72 ± 0.15	1.68 ± 0.21	1.75 ± 0.29	1.76 ± 0.28
H3	1.47 ± 0.37	1.44 ± 0.14	1.44 ± 0.10	1.31 ± 0.10
H4	1.72 ± 0.14	1.38 ± 0.23	1.63 ± 0.22	1.57 ± 0.21
H5	1.64 ± 0.33	1.51 ± 0.38	1.67 ± 0.18	1.71 ± 0.19
H6		1.67 ± 0.22	1.84 ± 0.41	1.83 ± 0.32
H7		1.73 ± 0.29	1.71 ± 0.27	1.69 ± 0.24
H8			1.77 ± 0.18	1.41 ± 0.38
H9				1.58 ± 0.41
$\frac{T(i)}{R(i-3) \text{ to } R(i+3)}$	1.655	1.637	1.696	1.676
H1	1.65 ± 0.11	1.68 ± 0.32	1.64 ± 0.08	1.65 ± 0.12
H2	1.71 ± 0.29	1.62 ± 0.06	1.77 ± 0.31	1.77 ± 0.28
H3	1.65 ± 0.20	1.33 ± 0.04	1.50 ± 0.25	1.58 ± 0.47
H4	1.71 ± 0.20	1.47 ± 0.13	1.80 ± 0.48	1.74 ± 0.44
H5	1.58 ± 0.10	1.43 ± 0.13	1.68 ± 0.21	1.72 ± 0.24
H6		1.65 ± 0.13	1.78 ± 0.46	1.85 ± 0.42
H7		1.75 ± 0.43	1.65 ± 0.12	1.63 ± 0.12
H8			1.69 ± 0.07	1.52 ± 0.28
H9				1.68 ± 0.12
$\frac{T(i)}{R(i-5) \text{ to } R(i+5)}$	1.968	1.915	1.984	1.984
H1	1.85 ± 0.26	1.89 ± 0.46	1.76 ± 0.14	1.86 ± 0.24
H2	1.94 ± 0.49	1.72 ± 0.12	2.01 ± 0.42	1.98 ± 0.35
H3	2.29 ± 0.54	1.44 ± 0.07	2.20 ± 0.54	2.33 ± 0.66
H4	1.91 ± 0.29	1.89 ± 0.50	2.27 ± 0.32	2.29 ± 0.33
H5	1.81 ± 0.38	1.76 ± 0.41	1.84 ± 0.25	1.84 ± 0.19
H6		1.82 ± 0.22	1.96 ± 0.38	2.06 ± 0.38
H7		2.00 ± 0.67	1.84 ± 0.28	1.79 ± 0.22
H8			1.85 ± 0.11	2.28 ± 0.08
H9				1.89 ± 0.29
$\frac{T(i)}{R(i-7) \text{ to } R(i+7)}$	2.290	2.228	2.306	2.319

H1	2.18 ± 0.68	1.96 ± 0.41	1.91 ± 0.26	1.95 ± 0.24
H2	2.24 ± 0.88	1.84 ± 0.31	2.25 ± 0.49	2.20 ± 0.43
H3	2.59 ± 0.30	1.66 ± 0.03	2.96 ± 0.63	3.19 ± 0.68
H4	2.14 ± 0.59	2.39 ± 0.55	3.27 ± 0.93	3.81 ± 1.20
H5	2.06 ± 0.70	2.08 ± 0.27	1.96 ± 0.25	2.04 ± 0.44
H6		2.02 ± 0.39	2.27 ± 0.54	2.51 ± 0.99
H7		2.40 ± 1.06	2.10 ± 0.50	1.99 ± 0.39
H8			1.97 ± 0.26	3.28 ± 0.17
H9				2.06 ± 0.38

Angle phase diagrams

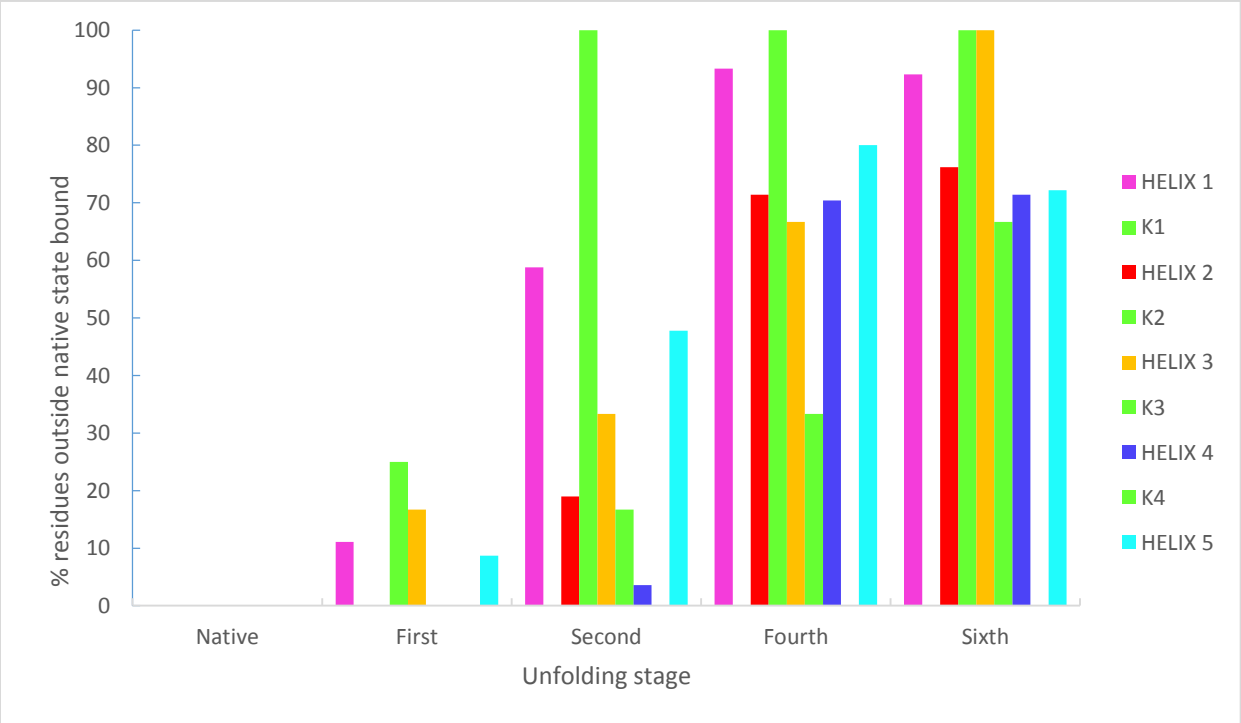
As noted above, for each residue i of each protein there is a specific, distinctive set of angles $[\alpha_i , \beta_i , \gamma_i]$ defining its relation to the iron atom. See Fig. 1. Relative to that atom, the couple $[\gamma_i , \alpha_i]$ specifies the angular orientation of terminal residues in each segment to the polypeptide backbone. The couple $[\beta_i,\alpha_i]$ specifies the angular orientation of the terminal residues in terms of the overall angle β_i between terminal residues and the angle α_i between $R(i-1)$ and $R(i-1 \text{ to } i+1)$. As the protein unfolds from its native state, the angle β_i increases and the angle α_i systematically decreases. In the couple $[\alpha_i , \gamma_i]$, both angles decrease as the protein unfolds.

We now display angle phase diagrams $[\gamma_i , \alpha_i]$ and $[\beta_i , \alpha_i]$ for the native and sixth extended states sequence of n residues in cyt c-b_{562} [Fig. 3] and h-Cygb [Fig. 4]. Figs. SI-1 and SI-2 display $[\gamma_i , \alpha_i]$ and $[\beta_i , \alpha_i]$ for the middle stages of unfolding. The solid black lines forming the triangles in these figures delimit the region $[\gamma_i , \alpha_i]$ and $[\beta_i , \alpha_i]$ defining the *native* state. As before [32], we adopt color coding to identify and distinguish the helical and non- helical regions of each protein: for cyt c-b_{562}, the first two helices from the N-terminal end are in magenta and red; the first two helices interior to the C-terminal end are in cyan and blue; the interior helix adjacent to the red helix is in gold; and the interior helix adjacent to the blue helix is in violet.

For h-Cygb with nine helices, the next "pairing" of interior helices is coded in yellow and gray, with the centermost helix in brown. Using this convention, similarities and differences between corresponding helices in the two proteins can be seen at a glance.

In following the unfolding of the protein, the departure of each (color coded) helical and non-helical region from the angular phase diagram specifying the native state (delimited by the boundary in black) can be tracked. See Tables 2-5. The results displayed in these tables will be discussed later.

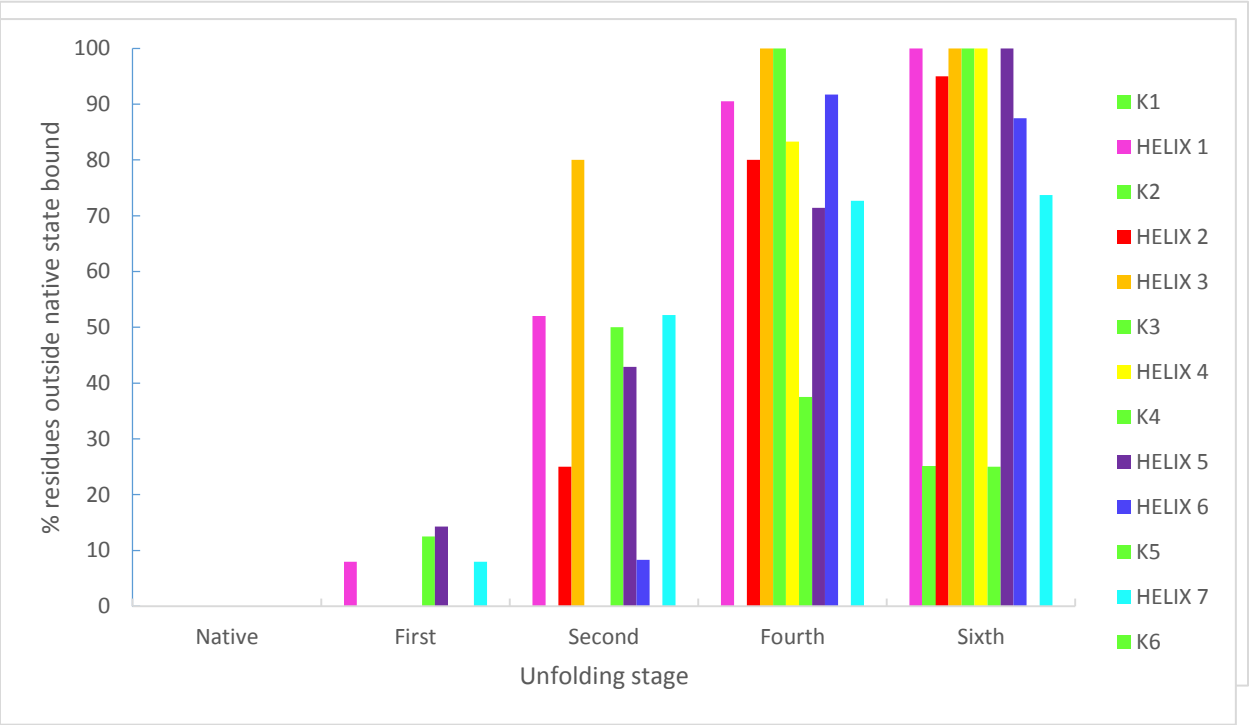
Table 2(a). Percent departure of helical and non-helical regions from native state



[β, α] domain: cyt c-b_{562}

Table 2(b). Percent departure of helical and non-helical regions from native state
[γ, α] domain: cyt c-b_{562}

Table 3(c). Percent departure of helical and non-helical regions from native state



[β, α] domain: cyt c'

Table 3(d). Percent departure of helical and non-helical regions from native state
[γ, α] domain: cyt c'

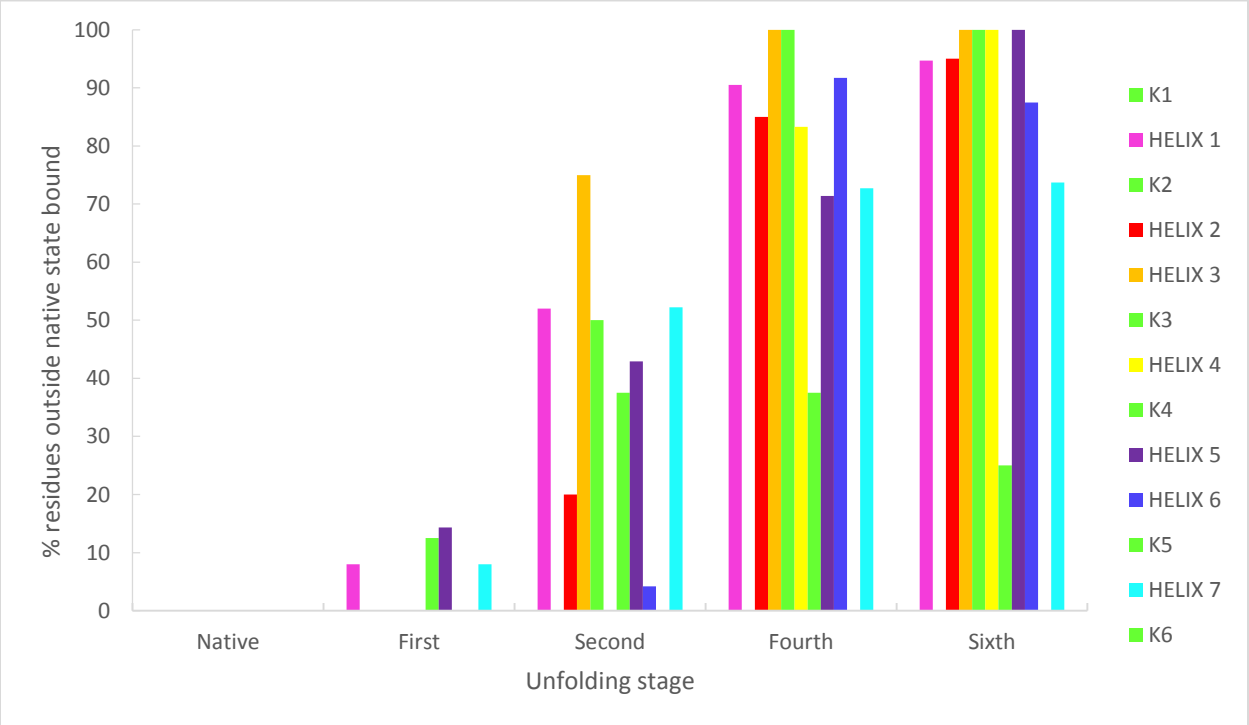


Table 4(e). Percent departure of helical and non-helical regions from native state
[β, α] domain: sw-Mb

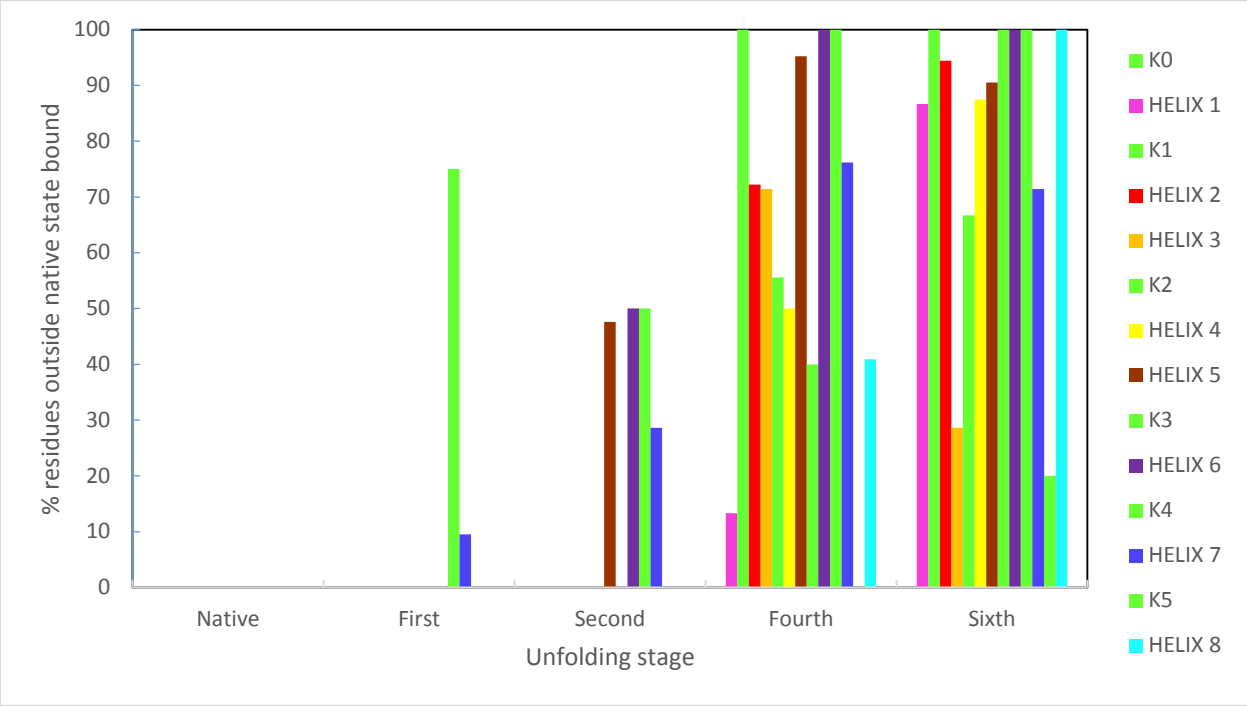


Table 4(f). Percent departure of helical and non-helical regions from native state
[γ, α] domain: sw-Mb

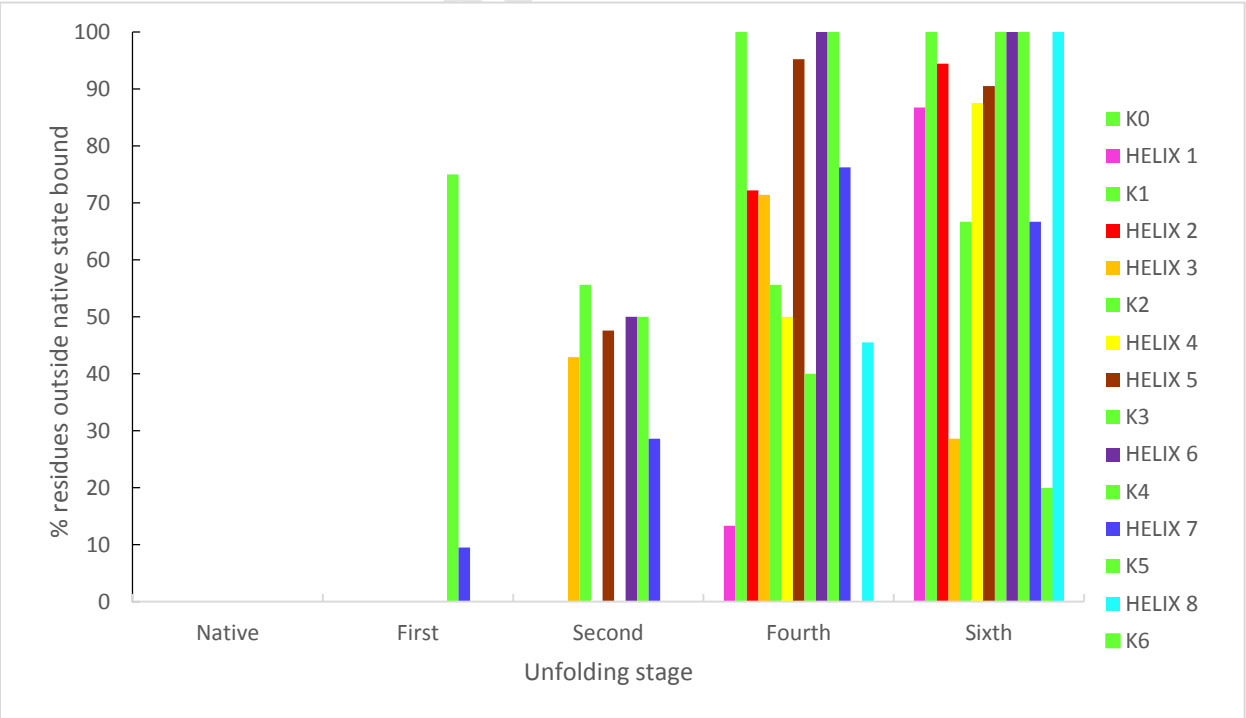


Table 5(g). Percent departure of helical and non-helical regions from native state [β, α] domain: h-Cygb

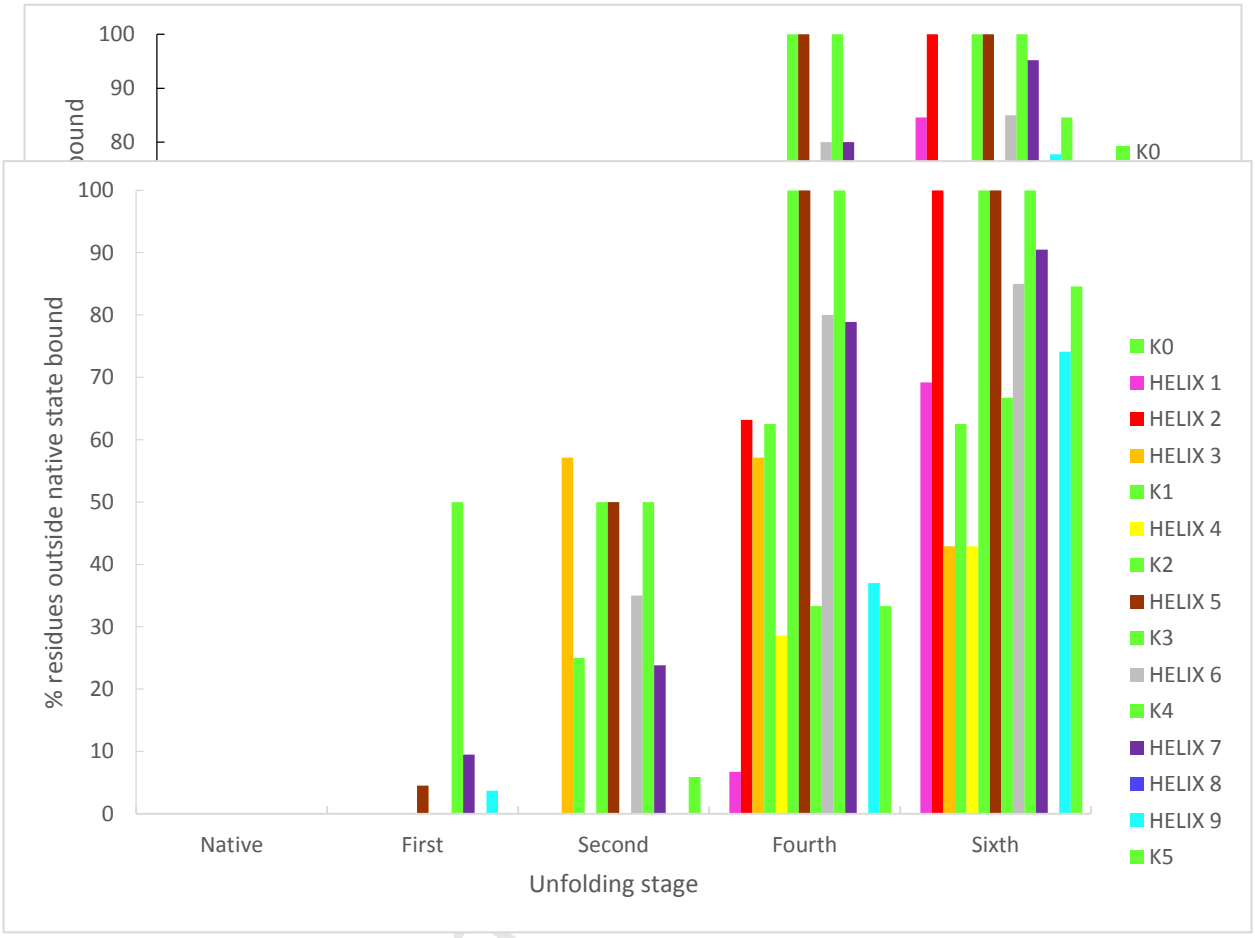
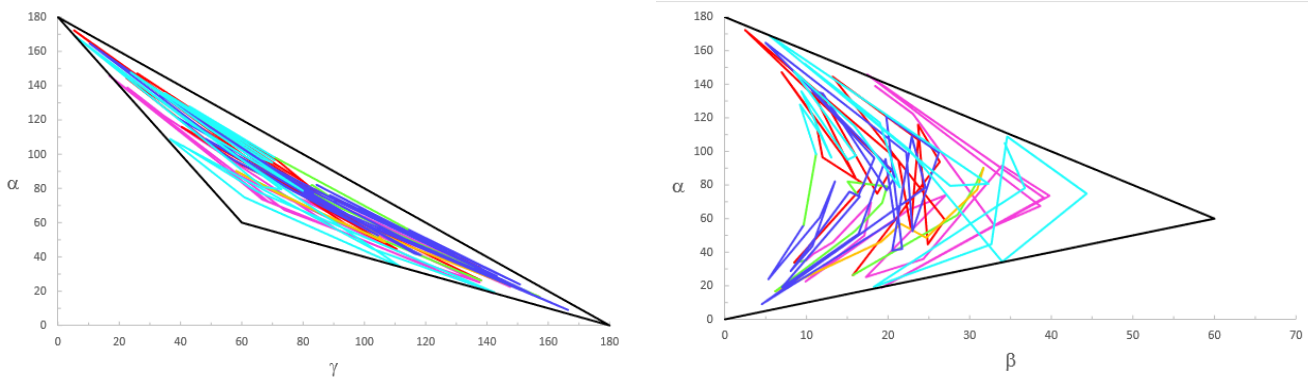


Table 5(h) Percent departure of helical and non-helical regions from native state [γ, α] domain: h-Cygb

Figure 3. Angle phase diagrams for cyt c-b_{562}, (*top*): {γ vs α} and {β vs α} native states, (*bottom*) : {γ vs α} and {β vs α} sixth extended states.



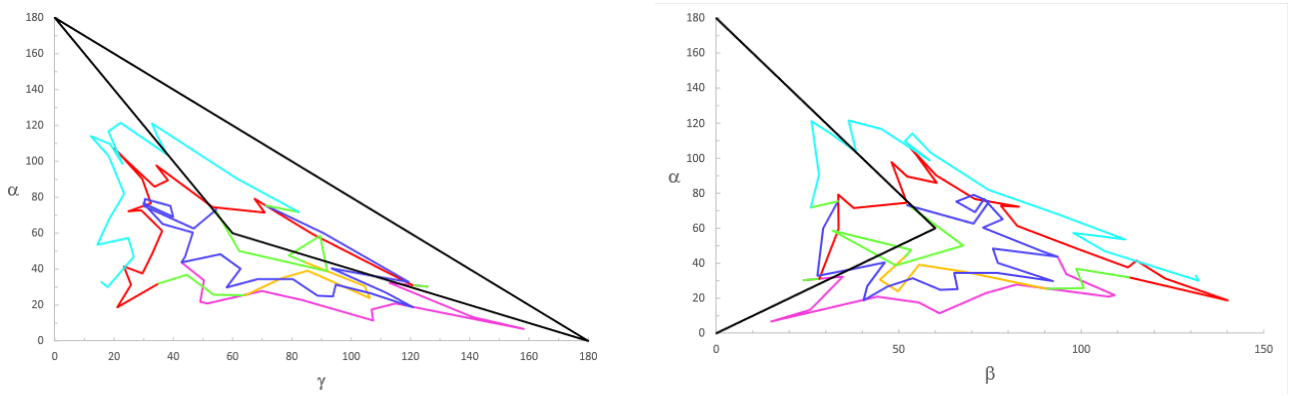
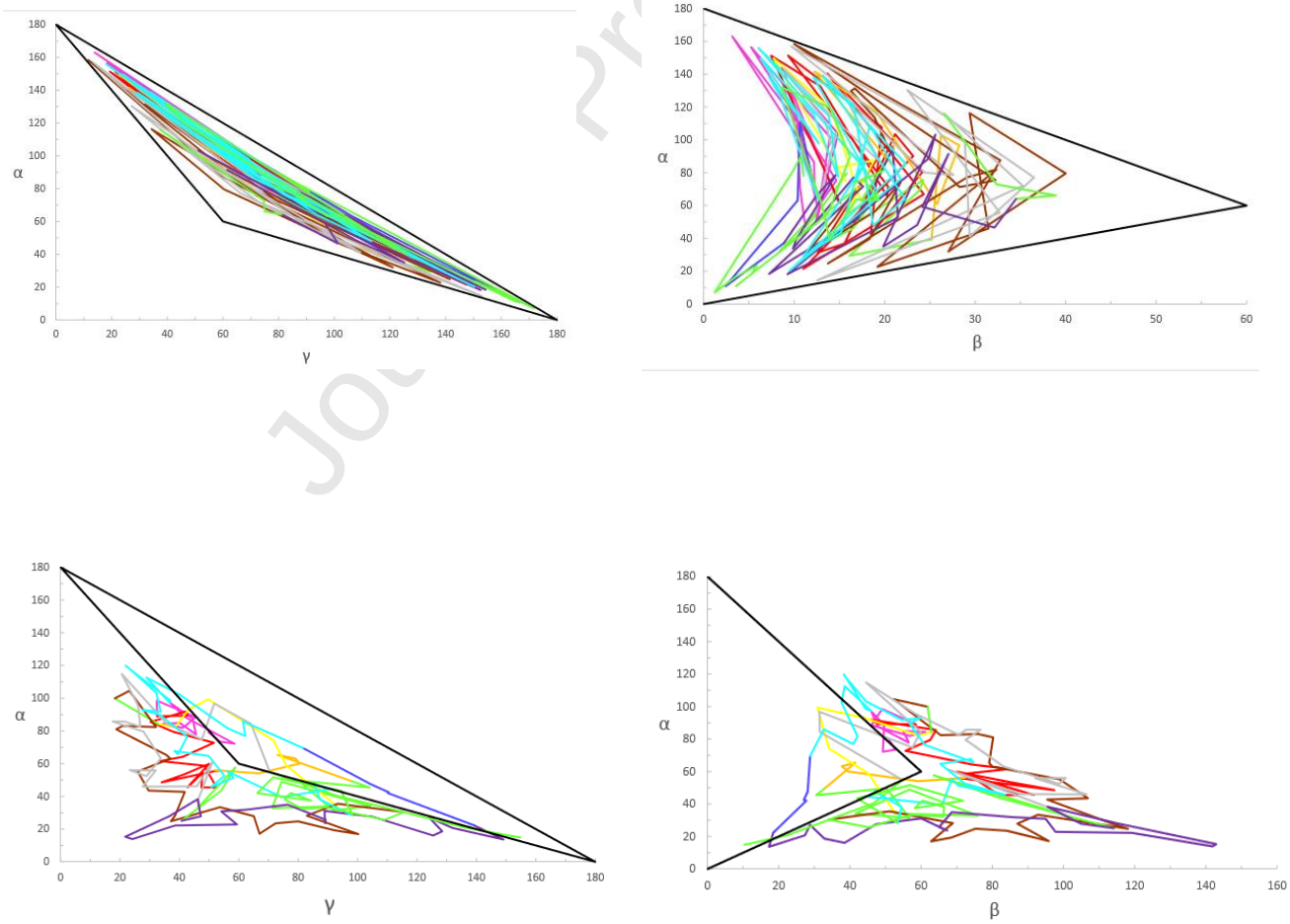


Figure 4. Angle phase diagrams for h-Cygb, (*top*): {γ vs α} and {β vs α} native states and (*bottom*) : {γ vs α} and {β vs α} sixth extended states.



Funnel diagrams

To visualize the sequential unfolding of a protein, we use the information encoded in the angular phase diagrams presented in the previous section. In our formulation, β_i in the triad [α_i, β_i, γ_i] is the angle between terminal α-carbons in a given segment, with the iron atom taken as the coordinate system origin. As the protein unfolds from its native state, angle β_i increases and angles [γ_i , α_i] systematically decrease.

Consider first the angular phase portraits displayed in plots of [γ_i , α_i] versus residue number *i*, starting from the native state, and progressing stepwise through the first, second, fourth and sixth stage of unfolding. We construct a d=3 representation, where [γ_i , α_i] is displayed as a function of the stage of unfolding (for cyt c-b_{562}, see Figure 5); and we find that the (d=2) areal profile [γ_i , α_i] describing the native state is the largest, with the areal profiles [γ_i , α_i] decreasing as the protein unfolds.

Also of interest are plots of [β_i, α_i] versus residue number *i* displayed as a function of the state of unfolding (for cyt c-b_{562} and h-Cygb, see Fig. 5; and Figs. SI-3 and SI-4 for cyt c' and sw-Mb). In these representations, the funnel, which is largest in the sixth stage of unfolding, becomes progressively smaller approaching the native state. Notice that the cross sections for the first and second unfolded state are very similar, with greater differences displayed as the system evolves to the fourth and sixth extended state. This behavior is entirely consistent with that uncovered in our study of the spatial signature.

Discussion

In each of four helical iron proteins, cyt c-b_{562}, cyt c', sw-Mb and h-Cygb, the first unfolded state is specified by a planar configuration in which the triplet $[i-2, i-1, i]$ is annexed to the triplet $[i, i+1,\ i+2]$. Having characterized the geometry of each triplet module, we focus on the central residue i in this five-residue segment, and calculate the distances $R(i-2)$ between Fe and the α-carbon of the left-most residue, $R(i+2)$between Fe and the α- carbon of the right-most residue, and the distance $R(i-2)$ to $R(i+2)$, between the terminal residues in this segment $i-2$ and $i+2$. As for the triplet, we then calculate relative to the Fe atom the angles between $R(i-2)$ and $R(i-2)$ to $R(i+2)$, $R(i-2)$ and $R(i+2)$, and $R(i-2)$ and $R(i+2)$ and $R(i+2)$, denoted α_i, β_i, γ_i, respectively.

Using the above data, we calculate first the displacement f_i relative to the Fe atom of the α-carbon of residue i from the native state to the first unfolded state. We then calculate $[\alpha_i, \beta_i, \gamma_i]$. This program is carried out for each residue of a given protein (except, in this case, the two residues at the C-terminal and N-terminal ends of the polypeptide chain).

One turn of an α-helix involves 3.6 residues; hence, in the first unfolded state, constructing the planar juxtaposition of two triplets, a five-residue segment, one H-bond is broken (or significantly weakened). The second unfolded state involves the planar juxtaposition of three planar triplets, a seven-residue segment. Since two turns of an α-helix require 7.2 residues, we expect (and find) that the first and second unfolded states should have quantitatively similar values of f_i and $[\alpha_i, \beta_i, \gamma_i]$. See Table 1 and Figs. 1 and 5 in [31] and Fig. 2.

The situation changes qualitatively and quantitatively when we calculate the spatial angular signatures for the fourth and sixth unfolded states. The fourth unfolded state involves 11

residues. There are 3 - 4 H-bonds per turn of α-helix. Three turns of an α-helix require 10.8 residues. In the fourth stage of unfolding, which involves 11 residues, three H-bonds are broken. Four turns of an α-helix require 14.4 residues. In the sixth unfolded state, which involves 15 residues, four H-bonds are broken. In these two unfolded states, we expect (and find) that the values of f_i and [α_i, β_i, γ_i] are quantitatively different from the first two unfolded states.

As noted earlier, the ratio of the elongation $T(i)$ of the polypeptide chain owing, for example, to the disruption of H-bonds and the distance between terminal residues in a given segment in the native state can be correlated exactly with the displacement relative to the Fe atom of the α-carbon of the central residue *i* from the native state. For example, in the first unfolded state

$$f_i = \frac{T(i)}{R(i-2) \text{ to } R(i+2)}$$

This is a general result and the correspondence can be established rigorously for residues in both helical and turning regions for any stage of unfolding. Of particular interest are the summaries of data on angle phase diagrams in Tables 2-5. The following observations follow from examination of the data in these tables.

Tables 2(a) to 5(h) display unfolding patterns of the four investigated proteins: cyt c-b_{562} (PDB ID: 2bc5) [33], cyt c' (PDB ID: 1mqv) [34], sw-Mb (PDB ID: 5yce) [35] and h-Cygb (PDB ID: 2dc3) [36]. The plots quantify the percent departure of helical and non-helical regions from the native state domains [β,α]and [γ,α] (Tables 2-5) for each helical and turning region of the protein as it unfolds stepwise from the native state through its first, second, fourth and sixth stages of unfolding in the [β,α] and [γ,α] domains. For each stage, the C-terminal end is on the right. The color code is specified in the text.

We focus here on the similarities and differences between sw-Mb and h-Cygb. The protein sw-Mb has 151 residues, eight helices (H1-H8) and seven non-helical (turning) regions (K0- K6). Human cytoglobin consists of 188 residues, nine helices (H1-H9) and six non-helical regions (K0-K5). Comparing both angular domains we find that the departure of five out of the nine helices and two out of the six non-helical regions in h-Cygb is ≥ 84 %. For sw-Mb six out of eight helices and three out of seven non-helical regions achieve the same percentage. In the [β,α] domain for h-Cygb, the first helix that departs from the native state, angular domain is H7 (quantitatively, a percentage of 9.5%). Subsequently, about 50% of H3 and H5 unfolds more rapidly by the second stage. By the sixth stage, helices 2, 5 and 7 show the greatest departure.

The same trends are observed in the [γ,α] domain for h-Cygb with the exception that the departure of H1 is 85% by stage six in the [β,α] domain as compared to 69% in the [γ,α] domain. For non-helical (turning) regions, K4 is the first region that unfolds in the first stage in both domains. K1, K2 and K4 are fully unfolded by the sixth stage followed by K5. K0 never departs the native state, angular domain. Considering next the unfolding of sw-Mb, we observe the following in both [γ,α] and [β,α] domains. The first segments that unfold are H7 and K4. In the fourth stage of evolution the following ordering is observed:

(K1, H6, K4) >H5>H7>H2>H3.

In the sixth stage of unfolding, we find

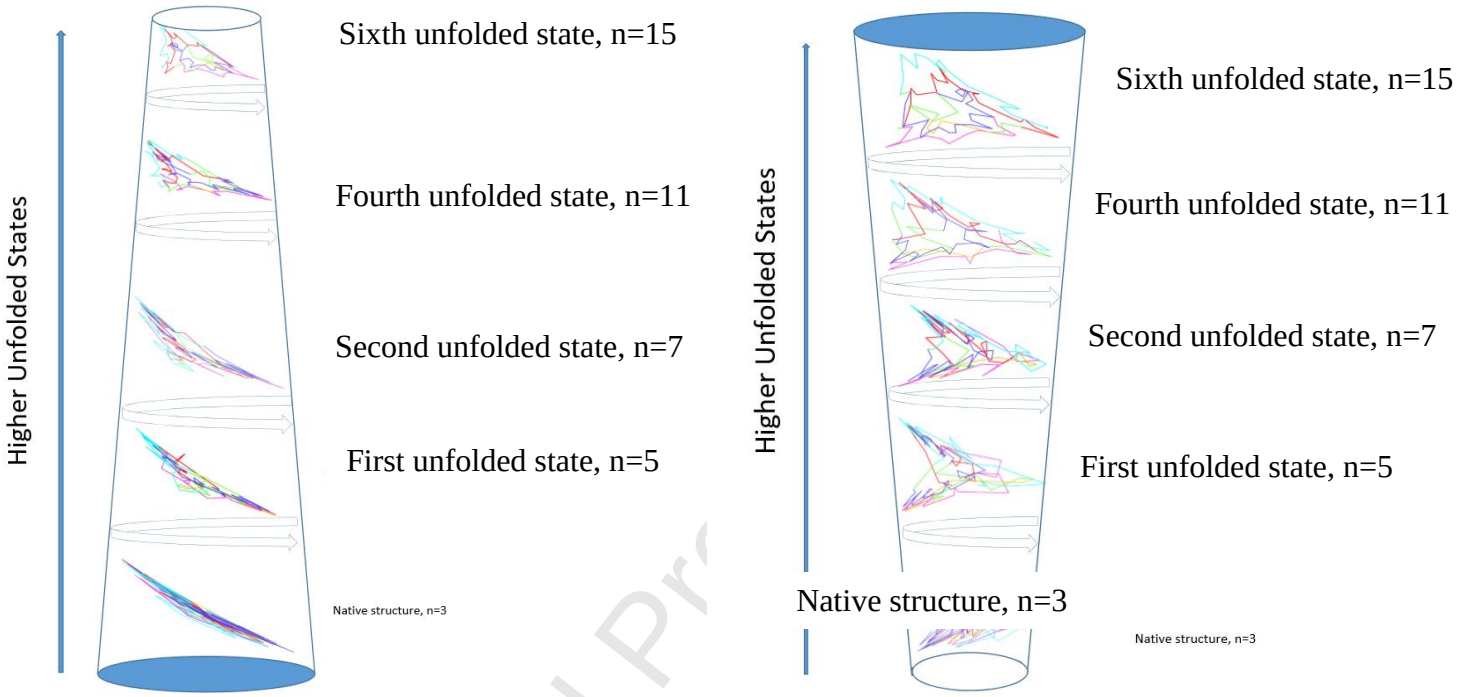
(K1, K3, H6, K4, H8)>H2>H5>H4>H1

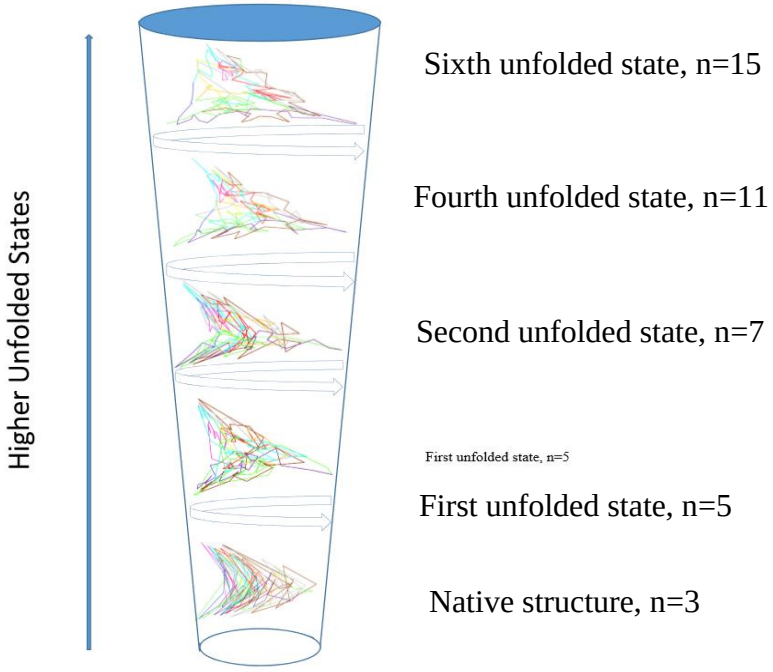
More than 80% of the helices H1, H2, H4, H5, H6 and H8 have exited the native state domain by the sixth stage while K1, K3 and K4 are fully unfolded at that stage. In contrast to h-Cygb, H2,

K2, H5, H6, K4, H7 and K5 of sw-Mb have completely exited (100%) from the native state domain, whereas K0 and K6 never depart the native state areal domain.

Funneled angle landscapes for cyt c-b_{562} and h-Cygb are in Figure 5; and those for cyt c' and sw-Mb are in the SI. As noted earlier, the areal profiles are very similar, qualitatively and quantitatively, for the first two stages of unfolding, with differences from the native state becoming more pronounced in the fourth and sixth stages of unfolding. The converse of this conclusion is that when these proteins fold, there are relatively large differences when stages well displaced from the native state are considered; indeed, we suggest that an interior helix protects the cyt c-b_{562} heme in the higher unfolded states, thereby disfavoring misligation from N-donor sidechains such as that of His63. Notably, in accord with experimental work on helical landscapes [11], there is a steep cascade down the funnel in the near vicinity of the native state, here the first two stages of unfolding.

Figure 5. *Top*: Funneled angle landscape for cyt c-b_{562} (left: [γ, α] vs unfolding stage, right: [β, α] vs unfolding stage. *Bottom*: Funneled angle landscape for h-Cygb; [β, α] vs unfolding stage.





So, what have we learned? First, we suggest that our funneled angle landscapes are complementary to funneled energy landscapes. At one level, this must be the case, since the angle diagrams are constructed from crystallographic data alone, data that represent the globally optimized protein structure in which all hydrophobic interactions, H-bond interactions, and excluded volume effects are taken into account. A complete energy landscape must necessarily include the key features of our angle diagrams. We suggest that these two representations are self-consistent: note that Onuchic and Wolynes [3,37] found that the frustration ratio, T_f/T_g, for a folding landscape is $T_f/T_g \sim 1.6$. As reported in Table 1, the departure from the native state, as quantified by the ratio f_i (see Appendix 1), is calculated to be $f_i \sim 1.6$ for the first two unfolded states, and then increasing as unfolded states farther from the native state are considered. Importantly, this trend, both qualitatively and quantitatively, is displayed by all four iron proteins. We suggest that this near correspondence between frustration ratio T_f/T_g and the ratio f_i that gauges the departure from the native state may provide the linkage between the two types of funneled landscapes, one based on energy, the other on angles defining the changing conformational structure of a protein.

Another take-home lesson follows from the profiles displayed in Fig. 2 and the body of evidence presented in Table 1. The data in Table 1 show that the average value of the spatial metric calculated for all four proteins is about the same at each stage of unfolding, increasing systematically as the protein unfolds. This universal behavior might have been anticipated, but the quantitative similarity is quite striking.

At a more detailed level, we find (see Fig. 2) that individual helical regions are characterized by values of the spatial metric for *interior* residues that are sensibly the same. However, residues adjacent to turning regions can display significant deviation from the more-or-less uniform value calculated for interior residues. This qualitative observation can be quantified by considering the standard deviation for each helical region recorded in Table 1. Compare the standard deviations for the N-terminal helix, and the adjacent helix in each protein, respectively H1 and H2. In the *first two* stages of unfolding, the average of the two standard deviations for H1 and H2 is very similar for all four proteins. The same similarity is found on comparing the helical region at the C-terminal end and the immediately adjacent, interior helix.

Importantly, however, in the *third* stage of unfolding, this coherence begins to unravel, and definitely in the *fourth* stage of unfolding, the standard deviations for the two cytochromes are *much* larger than the two globins. In our geometrical approach, progressing from the first, to the second, fourth and sixth stages of unfolding involves correlations among 5-, 7-, 11- and 15-residues. Hence, when the protein begins to unfold, residues in a helical region which are adjacent to turning regions are more and more influenced by residues in the turning region. Taken together the evidence suggests that turning regions (and, in particular, residues in turning regions immediately adjacent to helical regions) play a greater role in influencing the structural stability of the helix-bundle cytochromes than for myoglobin and human cytoglobin.

The question we now address is the relationship between our approach and that developed in the important series of papers by Thirumalai and coworkers [26-29]. Thirumalai and Guo [26] employed a minimalist approach to design a four-helix bundle model of minimum energy at temperature T, defined as the native state. After taking into account sequences that can fold into compact four-helix bundles and constructing bundles with three types of residues: hydrophobic

(B), hydrophilic (L) and neutral (N), they built an energy function based on three potentials: non-bonded, bond-angle and dihedral-angle (the non-bonded potential was defined as the interaction of two residues separated by at least three bonds along the chain; the bond-angle potential is the bond-angle deformation energy among three successive residues; and the dihedral-angle potential involves three successive bonds). They then started with denatured states at high temperature and lowered the temperature to allow their structures to collapse to a native state.

In subsequent work [27-28], the group constructed more sophisticated potentials to generalize their minimal protein model and thereby describe interactions within proteins in more detail. Importantly, this coarse-grained model has had an impact on the field, as it has given us a global picture of folding (it is this emphasis that distinguishes their approach from ours, since we proceed from the lowest energy configuration defined by the native state, and then move stepwise to unfold the polypeptide chain).

Our approach is based on a fundamental modular unit, a triplet of three nearest-neighbor residues. The geometry of each triplet modular unit comprising the protein is determined from the crystallographic data, directly, with no approximations. The net result of the molecular interactions of the side chain of the central residue (*i*) in the triplet with the side chains of the two adjacent residues (*i*-1) and (*i*+1) is embedded in the crystallographic data, and represents a globally minimized free energy state.

Formally, a potential function could be defined to describe interactions between the side chain of residues *i* and *i*-1, residues *i* and *i*+1 and residues *i*-1 and *i*+1. A quantum-chemical theory would then be mobilized to calculate the energetics of these interactions from first principles. In our modest approach, we bypass this step and use crystallographic data directly to reveal the

exact geometrical consequences of potential interactions between and among the three residues comprising the triplet.

We note that whatever ab initio quantum mechanical theory is implemented to calculate the interactions between the side chains in the triplet must recover the geometrical configurations we extract from the globally minimized structure of the native protein as captured by the crystallographic data. It is the sequential change in these optimized geometrical configurations that we characterize quantitatively as the protein unfolds.

Acknowledgements We thank Devarajan (Dave) Thirumalai for very helpful comments. Work at Caltech was supported by the NIH (DK019038) and the Arnold and Mabel Beckman Foundation. Support at Pomona College was provided by the Howard Hughes Medical Institute Research Program and a Sontag Research Fellowship Award.

References

1. Saven JG, Wolynes, PG (1996) Local conformational signals and the statistical thermodynamics of collapsed helical proteins: J Mol Biol 257, 199-216.

2. Onuchic JN, Luthey-Schulten Z, Wolynes PG (1997) Theory of Protein Folding: The energy landscape perspective: Ann Rev Phys Chem 48, 545-600.

3. Wolynes, PG (2015) Evolution, energy landscapes and the paradoxes of protein folding: Biochimie 119, 218-230.

4. Lin X., Schafer NP, Lu W, Jin, SX, Chen M., Onuchic JN, Wolynes, PG (2019) Forging tools for refining predicted protein structures: Proc Natl Acad Sci USA 116,9400-9409.

5. Onuchic JN, Wolynes PG (2004) Theory of protein folding: Curr Opin Struct Biol 14, 70 –75.

6. Plotkin SS, Onuchic JN (2002) Understanding protein folding with energy landscape theory.Part I. Basic concepts: Q Rev Biophys 35, 111–167.

7. Plotkin SS, Onuchic JN (2002) Understanding protein folding with energy landscape theory. Part II. Quantitative aspects: Q Rev Biophys 35, 205–286.

8. Chiti F, Dodson CM (2017) Protein Misfolding, Amyloid Formation, and Human Disease: A Summary of Progress Over the Last Decade: Annu Rev Biochem 86, 27-68.

9. Chung HS, EatonWA (2018) Protein folding transition path times from single molecule FRET: Curr Opin Struct Biol 48, 30-39.

10. Weinkam P, Pletneva EV, Gray HB, Winkler JR, Wolynes PG (2009) Electrostatic effects on funneled landscapes and structural diversity in denatured protein ensembles: Proc Natl Acad Sci USA 106, 1796 –1801.

11. Kimura T, Lee JC, Gray HB, Winkler JR (2009) Folding energy landscape of cytochrome cb_{562}: Proc Natl Acad Sci USA 106, 7834 –7839.

12. Telford, J.R., Wittung-Stafshede, P., Gray, H.B., Winkler, J.R. (1998) Protein folding triggered by electron transfer: Acc. Chem. Res. 31, 755-763

13. Le Duff CS, Whittaker SB, Radford SE, Moore GR (2006) Characterization of the conformational properties of urea-unfolded Im7: Implications for the early stages of protein folding: J Mol Biol 364, 824 – 835.

14. Borgia A, Gianni S, Brunori M, Travaglini-Allocatelli C (2008) Fst folding kinetics and stabilization of apocytochrome c: FEBS Lett 582, 1003–1007. 1

15. Moore CD, Lecomte JTJ (1990) Structural properties of apocytochrome b5: Presence of a stable native core. Biochemistry 29, 1984 –1989.

16. Moore CD, Almisky ON, Lecomte JTJ (1991) Similarities in structure between holocytochrome b5 and apocytochrome b5: NMR studies of the histidine residues: Biochemistry 30, 8357– 8365.

17. Garcia P, et al. (2005) Effects of heme on the structure of the denatured state and folding kinetics of cytochrome b562: J Mol Biol 346, 331–344.

18. Fuentes EJ, Wand AJ (1998) Local stability and dynamics of apocytochrome b562 examined by the dependence of hydrogen exchange on hydrostatic pressure: Biochemistry 37,9877–9883.

19. Manyusa S, Whitford D (1999) Defining folding and unfolding reactions of apocytochrome b5 using equilibrium and kinetic fluorescence measurements: Biochemistry 38, 9533–9540.

20. Feng HQ, Takei J, Lipsitz R, Tjandra N, Bai YW (2003) Specific nonnative hydrophobic interactions in a hidden folding intermediate: Implications for protein folding: Biochemistry 42, 12461–12465.

21. Feng HQ, Vu ND, Zhou Z, Bai YW (2004) Structural examination of value analysis in protein folding: Biochemistry 4, 14325–14331.

22. Choy WY, Zhou Z, Bai YW, Kay LE (2005) An 15N NMR spin relaxation dispersion study of the folding of a pair of engineered mutants of apocytochrome b562: J Am Chem Soc 127, 5066 –5072.

23. Feng H, Zhou Z, Bai Y (2005) A protein folding pathway with multiple folding intermediates at atomic resolution: Proc Natl Acad Sci USA 102, 5026 –5031.

24. Zhou Z, Huang YZ, Bai YW (2005) An on-pathway hidden intermediate and the early rate-limiting transition state of Rd-apocytochrome b562 characterized by protein engineering: J Mol Biol 352, 757–764.

25. Wang T, et al. (2007) Probing the folding intermediate of Rd-apocytochrome b562 by protein engineering and infrared T-jump: Protein Sci 16, 1176 –1183.

26. Guo Z, Thirumalai D (1996) Kinetics and Thermodynamics of Folding of a de Novo Designated Four-helix Bundle Protein: J Mol Biol 263, 323-343.

27. Buchete N-V, Straub JE, Thirumalai D (2004) Development of novel statistical potentials for protein fold recognition: Curr Opin Struct Biol 14, 225-232.

28. Buchete N-V, Straub, JE, Thirumalai D (2003) Anisotropic coarse-grained statistical potentials improve the ability to identify nativelike protein structures: J Chem Phys 118, 7658-71.

29. Denesyuk NA, Thirumalai D (2015) How do metal ions direct ribozyme folding? Nat Chem 7, 793-801.

30. Kauzmann W (1959) Some factors in the interpretation of protein denaturation Adv Protein Chem 14, 1-63.

31. Kozak JJ, Gray HB (2019) Stereochemistry of residues in turning regions of helical proteins: J Biol Inorg Chem 24. 879-888.

32. Kozak JJ, Gray HB., Garza-López RA (2016) Cytochrome unfolding pathways from computational analysis of crystal structures: J Inorg Biochem 155, 44-55.

33. Faraone-Mennella J, Tezcan F A, Gray HB, Winkler JR (2006) Stability and Folding Kinetics of Structurally Characterized Cytochrome c-b(562): Biochem 45,10504-10511.

34. Lee JC, Engman KC, Tezcan FA, Gray HB, Winkler JR (2002) Structural Features of

Cytochrome c' Folding Intermediates Revealed by Fluorescence Energy-Transfer Kinetics: Proc Natl Acad Sci USA 99, 14778-14782.

35. Isogai Y, Imamura H, Nakai S, Sumi T, Takahashi KI, Nakagawa T, Tsuneshige A, Shirai T (2018) Tracing whale myoglobin evolution by resurrecting ancient proteins: Sci Rep 8, 16883-16883.

36. Makino M, Sugimoto H, Sawai H, Kawada N, Yoshizato K, Shiro, Y (2006) High-resolution structure of human cytoglobin: identification of extra N- and C-termini and a new dimerization mode: Acta Crystallogr Sect.D 62: 671-677.

37. Onuchic JN, Wolynes PG, Luthey-Schulten Z, Socci ND (1995) Toward an outline of the topography of a realistic protein-folding funnel: Proc Natl Acad Sci USA 92, 3626- 3630.

APPENDIX 1.

We now show that this expression for f_3 can be re-expressed exactly in an equivalent expression that is useful in interpreting results obtained in our analysis. We isolate the angular factor in f_3.

$FACTOR =$

$$\frac{\sin\left(\arccos\left[(R5^2 + R(1\ to\ 5)^2 - R1^2)\sqrt{2R5\ R(1\ to\ 5)}\right]\right)}{\sin\left(\arccos\left[(R1^2 + R5^2 - R(1\ to\ 5)^2)\sqrt{2\ R1\ R5}\right]\right)/R1}$$

Recalling that,

$$\sin\left[\arccos(x)\right] = \sqrt{(1-x^2)}$$

we re-express this $FACTOR$ as

$FACTOR =$

$$\frac{\sqrt{\left[\left(-R1^4 + \left(2(R(1\text{ to }5)^2 + R5^2)\right)R1^2 - \ (-R(1\text{ to }5) + R5)^2(R(1\text{ to }5) + R5)^2\ \right)/(R1^2R5^2)\ \right]}}{\sqrt{\left[\left(-R(1\text{ to }5)^4 + \left(2(R1^2 + R5^2)\right)R(1\text{ to }5)^2 - \ (R1 - R5)^2(R1 + R5)^2\ \right)/(R5^2R(1\text{ to }5)^2)\ \right]}}$$

On expanding and simplifying the algebra, we find, eventually, that:

$$FACTOR = \frac{1}{R(1\text{ to }5)}$$

Hence, formally,

$$f_3 = \frac{T_3}{R(1\text{ to }5)}$$

Inserting the values of T_3 and $R(1\text{ to }5)$ on the right-hand side we recover the numerical value 1.981.

We emphasize the importance of this (unexpected) analytic equivalence. As defined, f_3 is the displacement of residue 3 relative to Fe in the first stage of unfolding. T(i) is the distance

between the terminal α-carbons [*i*-2 to *i*+2] for a configuration in which the triplet [*i*-2, *i*-1, *i*] is annexed to the triplet [*i*, *i*+1, *i*+2]. Hence, f_3 is, as well, a measure of the elongation $\frac{T_3}{R(1 \text{ to } 5)}$ of the five-residue segment [*i*-2 to *i*+2] when the protein unfolds from the native state to the first unfolded state. It is remarkable that a signature f_3 defined to measure the "vertical" displacement of residue 3 from the Fe atom is exactly equal to the "horizontal" extension of the five-residue segment on unfolding.

Declaration of competing interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Graphical abstract

HIGHLIGHTS

- Geometrical model to study unfolding of four heme proteins
- Model quantifies the resiliency of the native state to steric perturbations
- Development of angular funnel diagrams
- Sequential unfolding of helical regions is predicted